Multi-parameter auto-models and their application

Cécile Hardouin, Jian-Feng Yao

Multi-parameter auto-models and their application

By Cécile HARDOUIN

SAMOS-MATISSE/Centre d'Economie de la Sorbonne/ Université Paris 1, 90 rue de Tolbiac, 75634 Paris Cedex 13, France. hardouin@univ-paris1.fr

And Jian-Feng YAO

IRMAR/Université de Rennes 1, Campus de Beaulieu, 35042 Rennes Cedex, France

jian-feng.yao@univ-rennes1.fr

Abstract

Motivated by the modelling of non Gaussian data or positively correlated data on a lattice, extensions of Besag's Markov random fields auto-models to exponential families with multi-dimensional parameters have been proposed recently. In this paper, we provide a multiple-parameter analog of Besag's one-dimensional result that gives the necessary form of the exponential families for the Markov random field's conditional distributions. We propose estimation of parameters by maximum pseudo-likelihood and give a proof for the consistency of the estimators for the multi-parameter auto-model. The methodology is illustrated with some examples, particularly the building of a cooperative system with beta conditional distributions.

Some key words: Auto-models; Multi-parameter exponential families; spatial cooperation; beta conditionals.

Corresponding author.

Cécile HARDOUIN
SAMOS-MATISSE/CES/Université Paris 1
90 rue de Tolbiac
75634 Paris Cedex
France
`Email: Cecile.Hardouin@univ-paris1.fr`

1 Introduction

We consider a random field $X = \{X_i, i \in S\}$ on a finite set of sites $S = \{1, \ldots, n\}$. For a *site* i, let $p_i(x_i|x^{(i)}) = p_i(x_i|x_j, j \neq i)$ be the *full conditional distribution*, that is the conditional density of X_i given the values of all the X_j's other than X_i, where we have used the notation $x^{(i)} = \{x_j : j \neq i\}$. An important approach in stochastic modelling consists of specifying the family of these conditional distributions $\{p_i(x_i|\cdot) : i \in S\}$, and then determining a joint distribution P of the random field that is compatible with this family (i.e., the p_i's are exactly the conditional distributions associated with P). The investigation of this problem dates back to the 1960's; see Whittle (1963) and Bartlett (1968). Let us recall that if the joint probability distribution P is positive wherever the marginal distributions are positive, the Hammersley-Clifford Theorem characterizes $\log P$ as being proportional to a sum of potentials deduced from a set of cliques. The milestone paper of Besag (1974) provides several key steps for the development of the subject, including a proof of the Hammersley-Clifford theorem, the introduction of auto-models, and popular estimation methods such as maximum pseudo-likelihood estimation. Related developments on conditionally specified models include a series of works by Arnold, Castillo, and Sarabia; see Arnold *et al.* (1999), Arnold *et al.* (2001) for complete references, even though their approach is not specifically suited to the Markov random-field framework.

In this paper, we focus on auto-models introduced by Besag (1974). This class of spatial models is constructed under two assumptions: first, the dependence between sites is pairwise and, secondly, the full conditionals belong to some exponential family. Special instances of auto-models include the so-called auto-logistic, auto-binomial, auto-Poisson, auto-exponential, auto-gamma and auto-normal schemes. However, these schemes have a major limitation: the sufficient statistic as well as the canonical parameter are *one-dimensional*. More precisely, the exponential families can involve more than one parameter, but both the sufficient statistic and the canonical parameter are one-dimensional: for instance, in the so-called auto-normal scheme, the conditional mean at each site i is expressed as a linear combination of the values at its neighbouring sites $\{x_j,\ j \neq i\}$, and the conditional variance is constant or depends only on the site i. Furthermore, integrability conditions have to be satisfied for the model to be well defined. As noticed by the author himself, several auto-models like the Poisson, exponential, and gamma schemes are of little practical interest, since the integrability condition ensures that only spatial *competition* between neighbouring sites can occur. However, mostly one would like to model spatial cooperation.

To overcome these drawbacks, significant effort has been put in by a number of authors. An extension of the first condition is proposed in Lee *et al.* (2001), where the pairwise dependence is replaced with a multiway dependence, but still with one parameter exponential families. Recently, a Markovian approach is proposed in Kaiser and Cressie (2000), where the commonly used positivity condition on the joint distribution is relaxed. To the best of our knowledge, the first attempt to extend the one-parameter exponential family set-up to a multi-parameter one was made by Cressie and Lele (1992), where the term *multi-parameter exponential family Markov random field models* was coined. Later, improvements were proposed in Kaiser and Cressie (2000), where a spatial model using beta conditional distributions, an exponential family with two-parameters, is analysed in detail. In Kaiser *et al.* (2002), Equation (7), the authors introduce a class of spatial models with general multi-parameter exponential family conditional distributions and raise the question of ensuring their compatibility with a joint distribution. A general answer to this question is the subject of this paper.

We give here the general parametrisation of multi-parameter auto-models, which is a new result. The main result of this paper, Theorem 1 in Section 2, determines the necessary form for multi-parameter exponential families in terms of the full conditionals. We provide a directly analogous result to that of Eq. (4.4) in Besag (1974).

Having established the general result, we examine several related problems. We begin with a simple illustration of an auto-model on two sites, which is interesting because we consider different state spaces. Then, in Section 3, we address the problem of building *cooperative* spatial models. In particular, we discuss auto-models with full conditionals that are beta distributed and we give explicit conditions on the parameters to ensure the integrability condition. These auto-models have the advantage of being able to exhibit spatial cooperation as well as spatial competition according to suitable choices of their parameter values. The results are more general than those of Kaiser and Cressie (2000) and Kaiser *et al.* (2002).

Next, in Section 4, the consistency of the pseudo-likelihood estimator in multi-parameter auto-models is established under quite general conditions. To give more insight into the effectiveness of this estimator, several simulation experiments are conducted for auto-models with beta conditional distributions and two different neighbourhood systems.

In Section 5, we give a discussion of our findings. Proofs of the theoretical results are gathered together in Section 6.

2 Multi-parameter auto-models

Recall the set of sites $S = \{1, \ldots, n\}$, and consider a measurable state space $(E, \mathcal{E}, m)$ (often a subset of $\mathbb{R}$). We let the configuration space $\Omega = E^S$ be equipped with the σ-algebra $\mathcal{E}^{\otimes S}$ and the product measure $\nu := m^{\otimes S}$. Although we consider $\Omega = E^S$, all the following results hold equally with a more general configuration space $\Omega = \prod_{i \in S} E_i$, where each individual space $(E_i, \mathcal{E}_i)$ is equipped with some measure m_i (we give such an example at the end of this section). A random field is specified by a probability distribution μ on Ω, and we will assume throughout the paper the positivity condition, namely, μ has an everywhere positive density P with respect to ν. Consequently we can write

$$\mu(dx) = P(x)\nu(dx) , \qquad P(x) = Z^{-1} \exp Q(x) , \tag{2·1}$$

where Z is a normalisation constant. From the Hammersley-Clifford Theorem, the energy function $Q(x)$ is a sum of potentials G defined on the set of underlying cliques. Moreover, the positivity condition implies that at each site i, the conditional distribution of $(X_i | X_j = x_j, j \neq i)$ has a density $p_i(x_i | x^{(i)})$ with respect to $m(dx_i)$ that is itself everywhere positive.

The two basic assumptions are as follows.

[B1] The dependence between the sites is pairwise-only, that is,

$$Q(x) = \sum_{i \in S} G_i(x_i) + \sum_{\{i,j\}} G_{ij}(x_i, x_j) .$$

We fix a *reference configuration* $\tau = (\tau_i) \in \Omega$. In most cases, $\tau = (0, \ldots, 0)$, but the choice of this reference configuration is arbitrary (Guyon (1995), Kaiser and Cressie (2000)). In the case of the beta conditional distributions in Section 3, $E = (0, 1)$ and we take $\tau = (\frac{1}{2}, \ldots, \frac{1}{2})$. The following notation is useful. If $x \in \Omega$, for each i we denote $\tau_i x$ to be the realisation deduced from x replacing x_i with τ_i.

Next, the potential functions are uniquely determined if we assume for all i, j, and $x \in \Omega$ that

$$G_{ij}(\tau_i, x_j) = G_{ij}(x_i, \tau_j) = G_i(\tau_i) = 0 . \tag{2·2}$$

Note that if this condition were not naturally satisfied, we may substitute for $G_{ij}(x_i, x_j)$,

$$G_{ij}(x_i, x_j) - G_{ij}(\tau_i, x_j) - G_{ij}(x_i, \tau_j) + G_{ij}(\tau_i, \tau_j) ,$$

and make a similar adjustment for $G_i(x_i)$. Thus, from (2·1), we have $Q(\tau) = 0$ and $Z^{-1} = P(\tau)$.

The second assumption generalises Besag's auto-models schemes of one-parameter exponential families to multi-parameter exponential families:

$$[\mathbf{B2}]: \quad \log p_i(x_i|x^{(i)}) = \langle A_i(x^{(i)}), B_i(x_i)\rangle + C_i(x_i) + D_i(x^{(i)}) \ , \quad A_i(x^{(i)}) \in \mathbb{R}^l, \quad B_i(x_i) \in \mathbb{R}^l.$$

The main result of the paper is the following theorem, which determines the necessary form of the local natural parameters $\{A_i(.)\}$ to ensure the compatibility of the family of full conditional distributions.

Theorem 1 *Assume that the two conditions **[B1]** and **[B2]** are satisfied with the normalisation $B_i(\tau_i) = C_i(\tau_i) = 0$ in **[B2]** . Furthermore, assume the following condition*

$$[\mathbf{C}]: \quad \textit{For all } \ i \in S, \ \mathrm{Span}\{B_i(x_i) : \ x_i \in E\} = \mathbb{R}^l.$$

Then, necessarily, the functions A_i take the form:

$$A_i(x^{(i)}) = \alpha_i + \sum_{j \neq i} \beta_{ij} B_j(x_j) \ ; \ i \in S, \tag{2·3}$$

where $\{\alpha_i \ : \ i \in S\}$ is a family of l-dimensional vectors, and $\{\beta_{ij} \ : \ i,j \in S, \ i \neq j\}$ a family of $l \times l$ matrices $\{\beta_{ij}\}$ satisfying $\beta_{ij}^T = \beta_{ji}$. Moreover, the potentials are given by

$$G_i(x_i) = \langle \alpha_i, B_i(x_i)\rangle + C_i(x_i) \ , \tag{2·4}$$

$$G_{ij}(x_i, x_j) = B_i^T(x_i)\beta_{ij}B_j(x_j) \ . \tag{2·5}$$

A model satisfying the assumptions of the theorem is called a *multi-parameter auto-model*. The additional condition [**C**] is not present in the one-parameter case, since it is automatically satisfied, because the B_i's are not identically zero. We shall see below that this condition is not restrictive and is easily satisfied in most examples.

Another important property of the model is that of symmetry. The general formulation given above does not impose any symmetry, and hence it can be useful for modelling random fields on arbitrary or oriented graphs. As an illustration, a simple auto-model is given at the end of this section on two sites that play an asymmetrical role. On the other hand, in the case of a spatially symmetrical random field, it is necessary that all the potentials $G_{ij}(x_i, x_j)$ are symmetric functions or, equivalently, that all the matrices β_{ij} are symmetric.

It is interesting at this point to compare the necessary form (2·3) and several existing forms proposed in Kaiser *et al.* (2002). It is not difficult to see that their three proposed forms, Eqs. (10), (11), and (12) of Kaiser *et al.* (2002), correspond respectively to the cases,

$$\beta_{ij} = \eta_{ij} \begin{pmatrix} 1 & \cdots & 1 \\ 1 & \vdots & 1 \\ 1 & \cdots & 1 \end{pmatrix} ; \quad \eta_{ij} = \eta_{ji} \in \mathbb{R} \ , \tag{2·6}$$

$$\beta_{ij} = \begin{pmatrix} \eta_{ij1} & \cdots & 0 \\ 0 & \ddots & 0 \\ 0 & \cdots & \eta_{ijl} \end{pmatrix} ; \quad \eta_{ijk} = \eta_{jik} \in \mathbb{R}, \ 1 \leq k \leq l \ , \tag{2·7}$$

and

$$\beta_{ij} = \eta_{ij} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} ; \quad \eta_{ij} = \eta_{ji} \in \mathbb{R}. \tag{2·8}$$

Even though these specific forms are useful in practice, the general result is given by Eqs. (2·4) and (2·5).

The following proposition is useful, giving a converse to Theorem 1. It also provides a practical way to choose the parameters for a well defined multi-parameter auto-model. Indeed, the only additional condition one must check in practice is that the energy function Q is *admissible* in the sense of the integrability condition:

$$\int_{\Omega} e^{Q(x)} \nu(dx) < \infty \ . \tag{2·9}$$

Proposition 1 *Assume that the energy function Q is defined by **[B1]** with potentials G_i, G_{ij} given in (2·4) and (2·5). Assume further that the integrability condition (2·9) holds. Then the family of conditional distributions $p_i(x_i|x^{(i)})$ belong to a multi-parameter exponential family given by **[B2]** whose natural parameters $A_i(x^{(i)})$ satisfy (2·3).*

We now give a simple example illustrating Theorem 1. Consider just two variables (X_1, X_2) such that the conditional distribution of X_1 given $X_2 = x_2$ is a gamma distribution, and X_2 given $X_1 = x_1$ is a Gaussian distribution. This example with $S = \{1, 2\}$ is interesting since the two state spaces are different, and the model is not symmetric. The reference configuration is $\tau = (1, 0)$. In other words, we have according to [**B2**]:

$$\log p_1(x_1|x_2) = \log f_{\theta_1(x_2)}(x_1) = \langle A_1(x_2), B_1(x_1) \rangle - D_1(x_2) \ ,$$
$$\log p_2(x_2|x_1) = \log g_{\theta_2(x_1)}(x_2) = \langle A_2(x_1), B_2(x_2) \rangle - D_2(x_1) \ .$$

Here $l = 2$ and B_i's in [**B2**] are $B_1(x) = (-x + 1, \log x)^T$ and $B_2(x) = (x, x^2)^T$.

Condition **[C]** is trivially satisfied here. Therefore, by Theorem 1, there exist two vectors $\alpha_1,\ \alpha_2$ of $\mathbb{R}^2$ and a 2×2 matrix β such that

$$A_1(x_2) = \alpha_1 + \beta B_2(x_2)\ , \quad A_2(x_1) = \alpha_2 + \beta^T B_1(x_1)$$

The joint density is $P(x_1, x_2) = P(\tau)\exp Q(x_1, x_2)$ with $Q(x_1, x_2) = \langle\alpha_1, B_1(x_1)\rangle + \langle\alpha_2, B_2(x_2)\rangle + B_1^T(x_1)\beta B_2(x_2)$.

If we do not consider the matrix β to be symmetric, and it does not have to be, the model contains 8 parameters. Explicit conditions on these parameters can be obtained straightforwardly from (2·9) to ensure admissibility of the energy function Q. In fact, this is a known result presented in Arnold *et al.* (1999), §4.8, but our derivation is simpler.

3 A special class of auto-models with beta conditionals

As pointed out in Besag (1974), several one-parameter auto-models necessarily imply *spatial competition* but not spatial cooperation between neighbouring sites. For instance, this is the case for the auto-exponential, the auto-Poisson and the auto-gamma schemes. This competitive behaviour is clearly inadequate for many spatial systems where neighbouring sites are indeed cooperative. A common way to get rid of this drawback is to transform the variables onto a bounded range. For instance a truncation or projection procedure could be used.

Another possible way to get cooperative auto-models is by using multi-parameter auto-models such as beta conditional distributions. Notice that the family of beta distributions offers a large variety of densities on a bounded interval $[a, b]$, which makes the auto-beta models a potentially important class of spatial models.

Consider the univariate beta density on $(0, 1)$ with parameters $p, q > 0$:

$$f_\theta(x) = \kappa(p, q)x^{p-1}(1-x)^{q-1} = \exp\{\langle\theta, B(x)\rangle - \psi(\theta)\}\ , 0 < x < 1,$$

where $\theta = (p-1, q-1)^T$, $B(x) = [\log(2x), \log(2(1-x))]^T$, $\psi(\theta) = (p+q-2)\log 2 + \log\kappa(p, q)$, and $\kappa(p, q) = \Gamma(p+q)/[\Gamma(p)\Gamma(q)]$. Throughout this section, we denote the two components of $B(x)$ by $u(x) = \log(2x)$ and $v(x) = \log[2(1-x)]$. Notice that $u(\frac{1}{2}) = v(\frac{1}{2}) = 0$.

We now consider a random field X with such beta conditional distributions and reference configuration $\tau = (\frac{1}{2}, \frac{1}{2}, ..., \frac{1}{2})$. Clearly, Condition **[C]** is satisfied. From Theorem 1, for $i, j \in S$ and $i \neq j$, there exist vectors $\alpha_i = (a_i, b_i)^T \in \mathbb{R}^2$ and 2×2 matrices $\beta_{ij} = \begin{pmatrix} c_{ij} & d_{ij} \\ f_{ij} & e_{ij} \end{pmatrix}$

satisfying $\beta_{ij} = \beta_{ji}^T$, such that

$$A_i(x^{(i)}) = \alpha_i + \sum_{j \neq i} \beta_{ij} B(x_j) = \alpha_i + \sum_{j \neq i} \beta_{ij} \begin{pmatrix} u(x_j) \\ v(x_j) \end{pmatrix} .$$

Furthermore, the energy function Q can be written as

$$Q(x_1, \ldots, x_n) = \sum_{i \in S} \langle \alpha_i, B(x_i) \rangle + \sum_{\{i,j\}} B^T(x_i) \beta_{ij} B(x_j) ,$$

and the reference configuration, $\tau = (\frac{1}{2}, \ldots, \frac{1}{2})$, satisfies $Q(\tau) = 0$.

This model is well defined if its energy function Q satisfies the integrability condition (2·9), something we now examine in detail. We first note that the natural parameters of the conditional beta distributions are given by

$$A_i(x^{(i)}) = \begin{pmatrix} A_{i,1}(x^{(i)}) \\ A_{i,2}(x^{(i)}) \end{pmatrix} = \begin{pmatrix} a_i + \sum_{j \neq i} \{c_{ij} u(x_j) + d_{ij} v(x_j)\} \\ b_i + \sum_{j \neq i} \{f_{ij} u(x_j) + e_{ij} v(x_j)\} \end{pmatrix} . \tag{3·1}$$

Since $p > 0$ and $q > 0$ defines the natural parameter space for the univariate beta distribution, it follows that for all i and all configurations $x^{(i)} \in (0,1)^{n-1}$,

$$1 + a_i + \sum_{j \neq i} \{c_{ij} u(x_j) + d_{ij} v(x_j)\} > 0 , \tag{3·2}$$

and

$$1 + b_i + \sum_{j \neq i} \{f_{ij} u(x_j) + e_{ij} v(x_j)\} > 0 . \tag{3·3}$$

We first consider the inequality (3·2) . If x_j tends to $0+$ or $1-$, it follows necessarily that $c_{ij} \leq 0$ and $d_{ij} \leq 0$. Consequently,

$$c_{ij} u(x_j) + d_{ij} v(x_j) = (c_{ij} + d_{ij}) \log 2 + c_{ij} \log(x_j) + d_{ij} \log(1 - x_j) \geq (c_{ij} + d_{ij}) \log 2.$$

Therefore, a sufficient condition for (3·2) is,

$$c_{ij} \leq 0, \quad d_{ij} \leq 0, \quad \text{and} \quad 1 + a_i > -(\log 2) \sum_{j \neq i} (c_{ij} + d_{ij}) . \tag{3·4}$$

Similarly, a sufficient condition for the second inequality (3·3) is

$$f_{ij} \leq 0, \quad e_{ij} \leq 0, \quad \text{and} \quad 1 + b_i > -(\log 2) \sum_{j \neq i} (f_{ij} + e_{ij}) . \tag{3·5}$$

Under these conditions, the family of beta *conditional* distributions $\{p_i(x_i | x^{(i)}) \, , i \in S\}$ is everywhere well defined.

As we now show in the following proposition, these conditions also ensure the admissibility of the energy function Q.

Proposition 2 *Assume that for all $i, j \in S$, the conditions (3·4) and (3·5) are satisfied. Then*

1. *The family of beta conditional distributions $\{p_i(x_i|x^{(i)})\ , i \in S\}$ is everywhere well defined.*

2. *The energy function Q satisfies the integrability condition (2·9).*

Consequently, the auto-model with beta conditional distributions is well defined by (3·1).

While the conditions (3·4) and (3·5) are already used in practice (see Kaiser and Cressie (2000) and Kaiser *et al.* (2002)), we are not aware of any published proof that they are sufficient for (2·9) and hence for the joint distribution to exist. Having well defined conditional distributions does not necessarily imply well defined joint distribution.

3·1 *Spatial cooperation versus spatial competition*

We now examine the spatial-competition and spatial-cooperation behaviour of the auto-beta model. At each site i, the mean of the conditional distribution $p_i(x_i|x^{(i)})$ is

$$\mathbb{E}(X_i|x^{(i)}) = \frac{1 + A_{i,1}(x^{(i)})}{2 + A_{i,1}(x^{(i)}) + A_{i,2}(x^{(i)})} \ .$$

The model is said to be *spatially cooperative* (respectively *competitive*) if, at each i, $\mathbb{E}(X_i|x^{(i)})$ is non-decreasing (respectively non-increasing) in each neighbouring value x_j, and is increasing (respectively decreasing) in at least one. Notice that $\mathbb{E}(X_i|x^{(i)})$ increases with $A_{i,1}(x^{(i)})$ and decreases with $A_{i,2}(x^{(i)})$.Therefore, the auto-beta model is spatially co-operative if, for all $i \neq j$, $c_{ij} = e_{ij} = 0$; and it is spatially competitive if, for all $i \neq j$, $d_{ij} = f_{ij} = 0$.

To conclude the discussion about the auto-beta models, we compare our results to those of Kaiser and Cressie (2000), specifically their eq. (16). In our notation, their auto-beta model corresponds to:

$$\beta_{ij} = -\eta_{ij} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \eta_{ij} \geq 0.$$

In other words, $c_{ij} = e_{ij} = 0$ and $d_{ij} = f_{ij} = -\eta_{ij}$, where $\eta_{ij} \geq 0$. This provides an auto-beta model with spatial cooperation as proved by Kaiser and Cressie (2000), but our results are more general. For example, the constraint $d_{ij} = f_{ij}$ is generally unnecessary, except in the case of a spatially symmetrical random field (see the remark following Theorem 1).

3·2 *A spatially cooperative model with a scheme that has four or eight nearest neighbours*

First consider the scheme with four nearest neighbours on a two-dimensional lattice, $S = [1, M] \times [1, N]$: each site $i \in S$ has four neighbours denoted as $\{i_e = i + (1, 0), i_w = i - (1, 0), i_n = i + (0, 1), i_s = i - (0, 1)\}$ (with obvious neighbour adjustments near the boundary). We assume translation invariance in the sense that the parameters are functions of the displacement between sites; we assume spatial symmetry, which implies $d_{ij} = f_{ij}$; we allow possible anisotropy between the horizontal and vertical directions; and we assume $c_{ij} = e_{ij} = 0$, in order to model spatial cooperation. Under all these conditions and from the result above, there exists a vector $\alpha = (a, b)$ and two 2×2 matrices,

$$\beta^{(k)} = d_k \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \ , \quad k = 1, 2, \tag{3·6}$$

such that for all i, vectors $\alpha_i = \alpha$, and for all $\{i, j\}$, matrices $\beta_{ij} = 0$ unless i and j are neighbours, in which case

$$\beta_{i,i_e} = \beta_{i_w,i} = \beta^{(1)}, \quad \beta_{i,i_n} = \beta_{i_s,i} = \beta^{(2)}.$$

The model involves 4 parameters (a, b, d_1, d_2). The integrability conditions (3·4) and (3·5) become

$$d_1 \leq 0, \quad d_2 \leq 0 \ ; \quad (1 + a) \wedge (1 + b) > -2(d_1 + d_2) \log 2 \ . \tag{3·7}$$

The conditional distributions are beta-distributed with natural parameters,

$$A_i(x^{(i)}) = \begin{pmatrix} a + d_1[v(x_{i_e}) + v(x_{i_w})] + d_2[v(x_{i_n}) + v(x_{i_s})] \\ b + d_1[u(x_{i_e}) + u(x_{i_w})] + d_2[u(x_{i_n}) + u(x_{i_s})] \end{pmatrix} \ . \tag{3·8}$$

We now enlarge the model to a scheme with eight nearest neighbours. Each site then has four more neighbours $\{i_{nw} = i - (1, 1),\ i_{ne} = i + (-1, 1),\ i_{sw} = i + (1, -1), i_{se} = i + (1, 1)\}$ (with neighbour adjustments near the boundary). Note that in this case, some cliques have three or four elements but we consider pairwise interactions only, as specified in the condition **[B1]**. We again assume translation invariance, spatial symmetry, and spatial anisotropy. We wish to model spatial cooperation. Consequently, there exists a vector $\alpha = (a, b)$ and four 2×2 matrices $\left\{\beta^{(k)}\right\}$ of the form of (3·6) with constants $\{d_k : \ k = 1, ..., 4\}$, such that for all i, vectors $\alpha_i = \alpha$, and for all $\{i, j\}$, matrices $\beta_{ij} = 0$ unless i and j are neighbours, in which case

$$\beta_{i,i_e} = \beta_{i_w,i} = \beta^{(1)},\ \beta_{i,i_n} = \beta_{i_s,i} = \beta^{(2)},\ \beta_{i,i_{nw}} = \beta_{i_{se},i} = \beta^{(3)},\ \beta_{i,i_{ne}} = \beta_{i_{sw},i} = \beta^{(4)}.$$

The model involves 6 parameters $(a, b, d_1, d_2, d_3, d_4)$. The integrability conditions (3·4) and (3·5) become

$$d_1, \ldots, d_4 \leq 0 \; ; \; (1+a) \wedge (1+b) > -2(\log 2) \sum_{k=1}^{4} d_k \; . \tag{3·9}$$

The conditional distributions are beta-distributed with natural parameters,

$$A_i(x^{(i)}) = \tag{3·10}$$
$$\begin{pmatrix} a + d_1[v(x_{i_e}) + v(x_{i_w})] + d_2[v(x_{i_n}) + v(x_{i_s})] + d_3[v(x_{i_{nw}}) + v(x_{i_{se}})] + d_4[v(x_{i_{ne}}) + v(x_{i_{sw}})] \\ b + d_1[u(x_{i_e}) + u(x_{i_w})] + d_2[u(x_{i_n}) + u(x_{i_s})] + d_3[u(x_{i_{nw}}) + u(x_{i_{se}})] + d_4[u(x_{i_{ne}}) + u(x_{i_{sw}})] \end{pmatrix}$$

We propose the method of maximum pseudo-likelihood to estimate the parameters of the multi-parameter auto-models.

4 Estimation for a multi-parameter auto-model

Parameter estimation for a Markov random field has been well studied. The method of maximum likelihood unfortunately needs computer-intensive approximations, since the likelihood function is known only up to a constant that involves the parameters. As a remedy, Besag (1974, 1977) proposed the method of maximum pseudo-likelihood. We refer the reader to Guyon (1995) for an account of theoretical investigations of the properties of maximum pseudo-likelihood estimators. We give below a result for the consistency of the maximum pseudo-likelihood estimator under assumptions of translation invariance, and this is followed by a simulation study to investigate the behaviour of the estimators in finite samples.

4·1 *Consistency of the pseudo-likelihood estimator on a lattice*

We now introduce specific notation for fields on the two-dimensional lattice $\mathbb{Z}^2$. The process is observed on a rectangle $\Lambda_n = [-n_1, n_1] \times [-n_2, n_2]$ where $n = (n_1, n_2)$ and $n_1 \wedge n_2 \to \infty$. As is usual in asymptotic theory, we assume translation invariance: that is, the neighbourhood relationship is defined through a *bounded set* V_0 of neighbours of the origin $(0, 0)$, such that the set of neighbours of an arbitrary site $i \in \mathbb{Z}^2$ is $V_i = i + V_0$, and the interaction coefficients that appear in the matrices $\{\beta_{ij}\}$ are possibly nonzero if and only if $j - i = u$, for some $u \in V_0$. Furthermore, translation invariance allows us to write the parameters of the multi-parameter auto-model as:

$$\theta = (\alpha, \; \beta_u, u \in V_0) \in \mathbb{R}^q,$$

where q denotes the dimension of θ.

For any subset $A \in \mathbb{Z}^2$, write $x_A = (x_i, i \in A)$ as the restriction of x on A. To emphasize the parameters θ, the family of local conditional distributions is written as:

$$\log p(x_i|x^{(i)}, \theta) = \log p(x_i|x_{V_i}, \theta) = \langle A(x_{V_i}, \theta), B(x_i)\rangle + C(x_i) + D(x_{V_i}, \theta) \ , \tag{4·1}$$

where

$$A(x_{V_i}, \theta) = \alpha + \sum_{u \in V_0} \beta_u B(x_{i+u}) \ . \tag{4·2}$$

To give the statement (and the proof) of the result, we need more notation and definitions. We suppose that for each θ, there exists a Gibbs distribution μ_θ on $E^{\mathbb{Z}^2}$ such that the conditional distributions 4·1 are those of μ_θ. Let the parameter space be

$$\Theta = \{\theta \in \mathbb{R}^q \ : \int_E \exp\left[\langle A(x_{V_0}, \theta), B(x_0)\rangle + C(x_0)\right] dm(x_0) < \infty \ \text{ for all } \ x_{V_0} \in E^{|V_0|}\}, \tag{4·3}$$

and let θ_0 be the true value of the parameter. Define $\mathcal{G}_s(\theta_0)$ to be the set of limiting Gibbs measures on $E^{\mathbb{Z}^2}$ that are translation invariant and compatible with the family of conditional distributions $\{p(x_i|x^{(i)}, \theta_0)\}$.

Denote $W_i = i \cup V_i$ and define

$$R(x_{W_i}, \theta) = \log[p(x_i|x_{V_i};\ \theta)/p(x_i|x_{V_i};\ \theta_0)] \ , \quad i \in \Lambda_n, \quad \theta \in \Theta. \tag{4·4}$$

Then the maximum pseudo-likelihood estimator introduced by Besag(1974, 1977) is defined by

$$\widehat{\theta}_n = \arg\min_{\theta \in \Theta} U_n(\theta) \quad ,$$

where

$$U_n(\theta) = -\frac{1}{|\Lambda_n|} \sum_{i \in \Lambda_n} R(x_{W_i}, \theta), \quad \theta \in \Theta. \tag{4·5}$$

and $|\Lambda_n| = \text{card}(\Lambda_n)$. Notice that the computation of $\widehat{\theta}_n$ does not need the value of θ_0 appeared in (4·4). We now show that $\widehat{\theta}_n$ is a consistent estimator under mild regularity conditions.

Theorem 2 *Assume:*

1. *The true (limiting) distribution belongs to* $\mathcal{G}_s(\theta_0)$.

2. *For all θ, $R(x_{W_0},\theta)$ is an integrable random variable in $L^1(\mu_{\theta_0})$.*

3. *The following identifiability condition holds: for any $\theta \in \Theta$, if $R(x_{W_0},\theta) = 0$ for almost all x_{W_0} (with respect to μ_{θ_0}), then necessarily $\theta = \theta_0$.*

Then, the pseudo-likelihood estimator $\widehat{\theta}_n$ converges to θ_0 almost surely (μ_{θ_0}), as $n_1 \wedge n_2 \to \infty$.

The assumptions made in Theorem 2 are natural. In particular, Assumption 3 ensures that the parametrisation $\theta \mapsto p(x_0|x_{V_0}\ ;\ \theta)$ is proper. For Assumption 1, as the dependence set V_0 is finite, it is well-known (see Sinai (1982)) that if the state space E is compact and the reference measure m finite, the set $\mathcal{G}_s(\theta_0)$ is not empty.

We give an application of this general theorem to the beta auto-models defined in Section 3·2 while assuming translation invariance, spatial symmetry, anisotropy and spatial cooperation. In relation to Equations (3·7) and (3·9), let us define

$$A_4 = \{(a,b,d_1,d_2)\ :\ \ d_1,d_2 \le 0;\ a\wedge b \ge -2(\log 2)(d_1+d_2)\}, \tag{4·6}$$

in the scheme with four nearest neighbours, and

$$A_8 = \{(a,b,d_1,d_2,d_3,d_4)\ :\ \ d_1,\dots,d_4 \le 0;\ a\wedge b \ge -2(\log 2)\sum_{k=1}^{4} d_k\}, \tag{4·7}$$

in the eight nearest neighbours system case.

Proposition 3 *Consider the auto-beta model of Section 3·2. Assume:*

1. *The parameter space Θ is a compact subset of A_4 or A_8 accordingly to the associated neighbours system.*

2. *The (true) limiting distribution μ_{θ_0} of the observations, defined on $E^{\mathbb{Z}^2}$, is translation invariant.*

3. *The following identifiability condition holds: for any $\theta \in \Theta$, if $R(x_{W_0},\theta) = 0$ for almost all x_{W_0} (with respect to μ_{θ_0}), then necessarily $\theta = \theta_0$.*

Then, the pseudo-likelihood estimator $\widehat{\theta}_n$ converges to θ_0 almost surely (μ_{θ_0}) as $n_1 \wedge n_2 \to \infty$.

4·2 *Simulation experiments*

We propose several simulation experiments to assess further the properties of the maximum pseudo-likelihood estimator. We consider the auto-models of Section 3·2, where we assumed translation invariance, anisotropy, spatial symmetry and spatial cooperation. Both the eight and the four nearest neighbours systems are examined with various lattice sizes. For each simulation, we ran a Gibbs sampler on a square lattice, in order to generate a sample from the auto-model (600 sweeps). Empirical estimates are computed from 1600 independent simulations: systematic errors from simulations are then of order $1600^{-\frac{1}{2}} = 0.025$.

4·2.1 Experiment with the eight nearest neighbours system

First we consider the model with eight nearest neighbours described by (3·9) and (3·10). We choose a set of parameter values that satisfy the integrability conditions (3·9) and allow spatial anisotropy between the four directions: $(a, b, d_1, d_2, d_3, d_4) = (12, 16, -1, -3, -0.5, -2)$. The lattice size is 64×64. Table 1 gives the bias averages and the standard deviations of the parameter estimates from 1600 independent runs. In this case, the maximum pseudo-likelihood method provides consistent estimators with however non-negligible standard deviations especially for small parameter values like d_1 or d_3.

Parameter	a	b	d_1	d_2	d_3	d_4
True values	12	16	-1	-3	-0.5	-2
Bias average	0.0263	0.0282	0.0018	-0.0117	-0.0077	0.0033
St. deviation	0.3503	0.4775	0.2956	0.2775	0.2457	0.2619

Table 1: Bias averages and standard deviations of the parameter estimates for the beta auto-model with eight nearest neighbours, on a 64×64 lattice, from 1600 independent runs.

4·2.2 Extended experiments with the four nearest neighbours system

Next we consider the model with four nearest neighbours as described in (3·7) and (3·8) with the four parameters (a, b, d_1, d_2). This model appears in Kaiser *et al.* (2002) as a latent process for the analysis of a real data set of diseased trees that involves a spatial hierarchical model. The authors impose the constraint $d_1 = d_2$, and propose Monte Carlo maximum likelihood estimation that results in the estimates $(a, b, d_1 = d_2) = (16.6, 18.9, -4.5)$. Notice that these values satisfy the integrability condition (3·7). Then we choose these values for our simulation experiments.

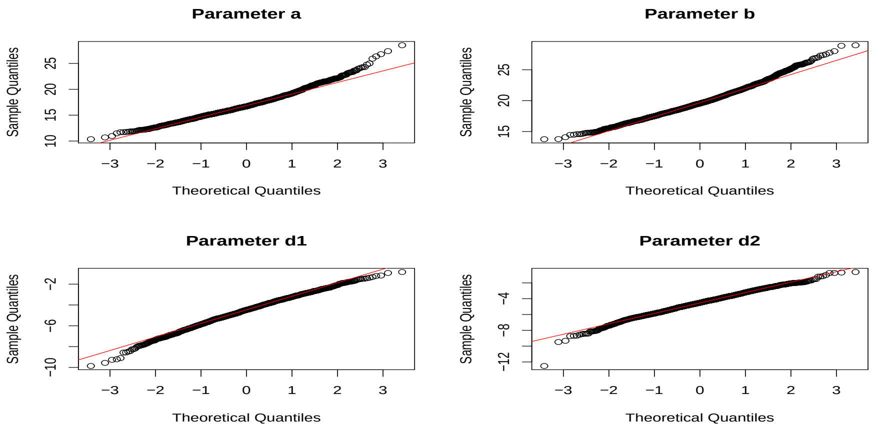
Here we are also interested in measuring empirically the convergence rate of the pseudo-likelihood estimators. Therefore, simulations are conducted on increasing lattice sizes: $n = 8 \times 8, 16 \times 16,\ 32 \times 32,\ 48 \times 48,\ 56 \times 56$, and 64×64. Average bias and standard deviations of the estimates are displayed in Table 2. We can see that the bias are quite large in the case of the smallest lattice size $n = 8^2$ and significantly reduced when $n = 64^2$.

Parameter		a	b	d_1	d_2
True values		16.6	18.9	-4.5	-4.5
8*8	Bias average	2.1640	4.5885	-0.4970	-0.0491
	St. deviation	6.6384	7.6895	3.8361	3.4149
16*16	Bias average	0.3574	0.8581	-0.0434	-0.0510
	St. deviation	2.3882	2.3416	1.3288	1.3376
32*32	Bias average	0.0260	0.2171	0.0273	-0.0066
	St. deviation	1.1393	1.1719	0.6372	0.6219
48*48	Bias average	0.0261	0.0852	-0.0210	0.0166
	St. deviation	0.7267	0.7263	0.4130	0.4359
56*56	Bias average	0.0319	0.0365	-0.0140	-0.0027
	St. deviation	0.6466	0.6547	0.3693	0.3619
64*64	Bias average	0.0307	0.0357	-0.0069	-0.0085
	St. deviation	0.5554	0.5713	0.3031	0.3179

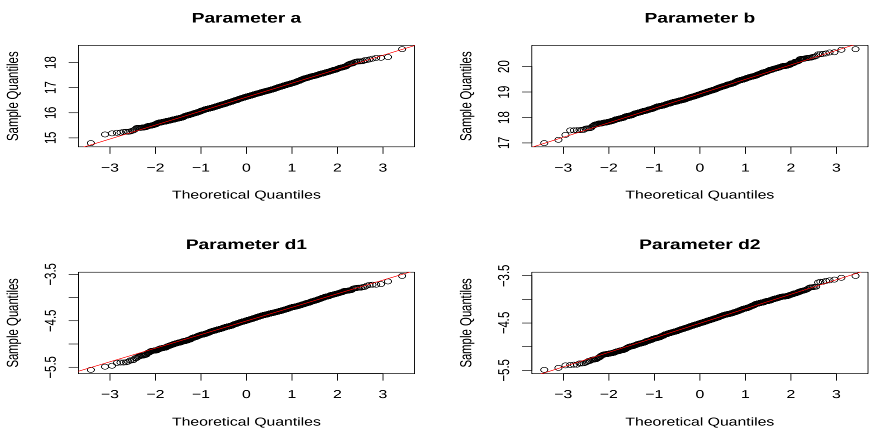
Table 2: Bias averages and standard deviations of the parameter estimates for the beta auto-model with four nearest neighbours, various lattice sizes and from 1600 independent runs in each case.

Next, to get insights on the sampling distributions, we examine Gaussian Q-Q plots of the estimates. Figure 1 below displays such plots for extreme sizes $n = 16^2$ and $n = 64^2$ (in the case of the smallest size $n = 8^2$, the plot shows a non Gaussian behaviour). For a size as small as $n = 16^2$, the plot is not so bad and the empirical distribution is close to a Gaussian distribution. Step by step, the Q-Q plots set right, and we finally get a "perfect" Gaussian approximation for $n = 64^2$.

Finally, we examine empirically the convergence rate of the estimates to their respective Gaussian distributions. Based on Table 2 and with the "bad" case $n = 8^2$ excluded, Figure 2 displays scatter-plots of the logarithms of the standard deviations versus $\log n$. Simple regression fits indicate a slope around $-\frac{1}{2}$ for all the four parameters, yielding a strong support for a root-n rate of their weak convergence to an asymptotic Gaussian distribution.



(a)



(b)

Figure 1: Gaussian Q-Q plots of the estimates from 1600 independent runs. a). Lattice size $n = 16^2$. b). Lattice size $n = 64^2$.

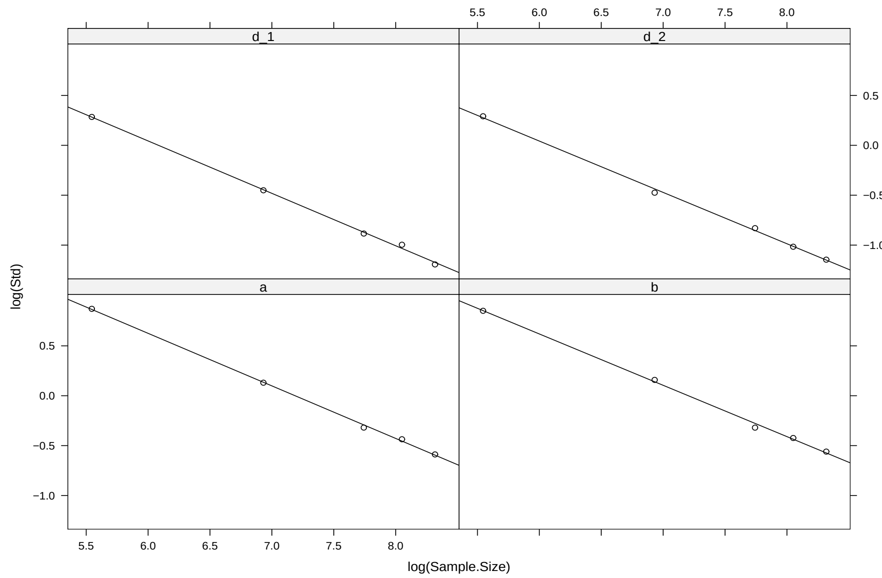


Figure 2: Standard deviations of estimates vs the sample size n (both in log-scale) with $n = 16^2, 32^2, 48^2, 56^2$ and 64^2. The slope of the regression line is -0.5264, -0.5142, -0.5250 and -0.5143 for parameters a, b, d_1 and d_2 respectively.

5 Discussion

In this paper we have proposed a directly analogous to Besag's auto-models to the situations where the local conditional distributions belong to some multi-parameter exponential families, such as the beta distribution. This allows one to model spatial cooperation (as well as spatial competition). Another interesting application is the modelling of mixed-state data where the distributions are mixtures of discrete and continuous components. Measurements can present continuous values during some periods and discrete values at other times, such as for daily rainfall time series. At a given site, there may be many zeros when the rain is absent, followed by periods with positive rainfall values (Allcroft and Glasbey (2003)); then, the state space becomes $E = \{0\} \cup (0, \infty)$. Any random variable X taking its values in E is called a *mixed-state* random variable, which we can define formally as follows: with probability $\gamma \in (0, 1)$, set $X = 0$, and with probability $1 - \gamma$, X is positive, continuous having a density belonging to a s-dimensional exponential family,

$$g_\xi(x) = H(\xi) \exp\langle \xi, T(x) \rangle, \ x > 0, \ \xi \in \mathbb{R}^s, \ T(x) \in \mathbb{R}^s.$$

Let $m(dx) = \delta_0(dx) + \lambda(dx)$ where δ_0 and λ are respectively the Dirac measure at 0 and the Lebesgue measure on $(0, \infty)$. Define the indicator function $\delta(x) = \mathbb{I}_{\{0\}}(x)$ and set

$\delta^*(x) = 1 - \delta(x),\ x \geq 0$. Then X has then the following density function (with respect to $m(dx)$):

$$f_\theta(x) = \gamma\delta(x) + (1-\gamma)\delta^*(x)g_\xi(x) = Z^{-1}(\theta)\exp\langle\theta, B(x)\rangle,\ x \geq 0,$$

where we have set

$$\theta = (\theta_1, \theta_2)^T = \left(\log\frac{(1-\gamma)H(\xi)}{\gamma}, \xi\right)^T \ , \quad B(x) = (\delta^*(x), T(x)^T)^T \ .$$

In other words, X belongs to a $(s+1)$-dimensional exponential family. When this formulation is applied to the conditional distributions on a lattice, we obtain a multi-parameter auto-model suitable for modelling data that are either zero or positive-valued. Theoretical results for these specific mixed-state auto-models will be studied elsewhere. Experimental application to motion measurements in video sequences can be found in (Bouthemy *et al.* (2006)).

On the other hand, the auto-model scheme will gather more power in applications if the assumed pairwise interactions are extended to more general multiway dependence. Another important question is to relax the positivity condition as proposed by Kaiser and Cressie (2000) in a general context. Finally, the simulations of Section 4 indicate that the maximum pseudo-likelihood estimators of the auto-models' parameters should be asymptotically normally distributed with a root-n convergence rate. It is clearly worth investigating theoretical studies to support such empirical evidence.

6 Proofs

Proof of Theorem 1

For each i, we have:

$$\begin{aligned} & Q(x) - Q(\tau_i x) \\ & = G_i(x_i) + \sum_{j:\ j\neq i} G_{ij}(x_i, x_j) = \log\frac{p_i(x_i|x^{(i)})}{p_i(\tau_i|x^{(i)})} \\ & = \langle A_i(x^{(i)}), B_i(x_i)\rangle + C_i(x_i)\ . \end{aligned}$$

By taking $x^{(i)} = \tau^{(i)} = \{\tau_j : j \neq i\}$, we obtain:

$$G_i(x_i) = \langle A_i(\tau^{(i)}), B_i(x_i)\rangle + C_i(x_i)\ . \tag{6·1}$$

Now, let us fix two indices $i \neq j$. For ease of exposition and without loss of generality, we may assume $i = 1$ and $j = 2$. The previous calculations also lead to

$$
\begin{aligned}
&Q(x_1, x_2, \tau_3, \ldots, \tau_n) - Q(\tau_1, x_2, \tau_3, \ldots, \tau_n) \\
&= G_1(x_1) + G_{12}(x_1, x_2) = \langle A_1(x_2, \tau_3, \ldots, \tau_n), B_1(x_1) \rangle + C_1(x_1) \ .
\end{aligned}
$$

Therefore,

$$
G_{12}(x_1, x_2) = \langle \mathcal{A}_1(x_2), B_1(x_1) \rangle \ ,
$$

where we have set

$$
\mathcal{A}_1(x_2) = A_1(x_2, \tau_3, \ldots, \tau_n) - A_1(\tau_2, \tau_3, \ldots, \tau_n) \ .
$$

In an analogous manner and switching the indices 1 and 2, we finally obtain for all $x_1, x_2 \in E$,

$$
G_{12}(x_1, x_2) = \langle \mathcal{A}_1(x_2), B_1(x_1) \rangle = G_{21}(x_2, x_1) = \langle \mathcal{A}_2(x_1), B_2(x_2) \rangle \ ;
$$

that is,

$$
B_1^T(x_1)\mathcal{A}_1(x_2) = \mathcal{A}_2^T(x_1) B_2(x_2) \ , \tag{6·2}
$$

where

$$
\mathcal{A}_2(x_1) = A_2(x_1, \tau_3, \ldots, \tau_n) - A_2(\tau_1, \tau_3, \ldots, \tau_n) \ .
$$

Next, Condition **[C]** for $i = 2$ means that there exist l elements of E, $\mathcal{Y} = \{y_1, ..., y_l\}$, such that the $l \times l$ matrix $B_2(\mathcal{Y}) = (B_2(y_1), \ldots, B_2(y_l))$ is invertible. We also write $\mathcal{A}_1(\mathcal{Y}) = (\mathcal{A}_1(y_1), \ldots, \mathcal{A}_1(y_l))$. Then, upon substituting $x_2 = y_j$, for $j = 1, \ldots, l$, into (6·2) ,

$$
\begin{aligned}
\mathcal{A}_2^T(x_1) B_2(\mathcal{Y}) &= [\mathcal{A}_2^T(x_1) B_2(y_1), \ldots, \mathcal{A}_2^T(x_1) B_2(y_l)] \\
&= [B_1^T(x_1)\mathcal{A}_1(y_1), \ldots, B_1^T(x_1)\mathcal{A}_1(y_l)] \\
&= B_1^T(x_1)\mathcal{A}_1(\mathcal{Y}).
\end{aligned}
$$

Therefore,

$$
\mathcal{A}_2^T(x_1) = B_1^T(x_1)\beta_{12}(\mathcal{Y}) \ , \qquad \text{where} \quad \beta_{12}(\mathcal{Y}) = \mathcal{A}_1(\mathcal{Y})[B_2(\mathcal{Y})]^{-1} \ .
$$

Consequently, G_{12} can be written as

$$
G_{12}(x_1, x_2) = B_1^T(x_1)\beta_{12}(\mathcal{Y}) B_2(x_2) \ .
$$

The left hand side of this equality does not depend on $\mathcal{Y}$, so $\beta_{12}(\mathcal{Y}) \equiv \beta_{12}$ is a constant matrix and we obtain,

$$G_{12}(x_1, x_2) = B_1^T(x_1)\beta_{12}B_2(x_2) \ . \tag{6·3}$$

By exchanging the indices, we also have $G_{21}(x_2, x_1) = B_2^T(x_2)\beta_{21}B_1(x_1)$. As $G_{12}(x_1, x_2) = G_{21}(x_2, x_1)$, for all $x_1,\ x_2$, we must have $\beta_{12} = \beta_{21}^T$.

Furthermore, $Q(x) - Q(\tau_1 x) = G_1(x_1) + \sum_{j\neq 1} G_{1j}(x_1, x_j)$. We use eqs. (6·1) and (6·3) in this expression and obtain,

$$A_1^T(\tau^{(1)})B_1(x_1) + C_1(x_1) + \sum_{j\neq 1} B_1^T(x_1)\beta_{1j}B_j(x_j) = A_1^T(x^{(1)})B_1(x_1) + C_1(x_1),$$

which is equivalent to

$$\alpha_1^T B_1(x_1) + \left(\sum_{j\neq 1} B_j^T(x_j)\beta_{j1}\right) B_1(x_1) = A_1^T(x^{(1)})B_1(x_1) \ .$$

That is,

$$\left[\alpha_1 + \sum_{j\neq 1} \beta_{1j}B_j(x_j) - A_1(x^{(1)})\right]^T B_1(x_1) = 0.$$

Hence, applying Condition [**C**] in the same manner as above, we obtain Equation (2·3) for $i = 1$.

Proof of Proposition 1

We have only to check that the conditional distributions of the field with potentials (2·4) and (2·5) are those given by [**B2**] and (2·3). This follows from:

$$\begin{aligned} Q(x) - Q(\tau_i x) &= G_i(x_i) + \sum_{j:j\neq i} G_{ij}(x_i, x_j) \\ &= \langle \alpha_i, B_i(x_i)\rangle + C_i(x_i) + \sum_{j\neq i} B_i^T(x_i)\beta_{ij}B_j(x_j) \\ &= \langle A_i(x^{(i)}), B_i(x_i)\rangle + C_i(x_i) = \log \frac{p_i(x_i|x^{(i)})}{p_i(\tau_i|x^{(i)})} \ . \end{aligned}$$

Proof of Proposition 2

We need only to prove the admissibility of Q. Let $\xi = (1,1)^T$ and $w(x) = B(x) - (\log 2)\xi = (\log x, \log(1-x))^T$. We have

$$B^T(x_i)\beta_{ij}B(x_j) = (\log 2)^2\xi^T\beta_{ij}\xi + (\log 2)[\xi^T\beta_{ij}w(x_j) + w^T(x_i)\beta_{ij}\xi] + w^T(x_i)\beta_{ij}w(x_j).$$

For all $i \neq j$, since $w^T(x_i)\beta_{ij}w(x_j) \leq 0$ and $\xi^T\beta_{ij}\xi \leq 0$, we have

$$B(x_i)^T\beta_{ij}B(x_j) \leq (\log 2)[\xi^T\beta_{ij}w(x_j) + w^T(x_i)\beta_{ij}\xi].$$

Therefore,

$$\begin{aligned}
Q(x_1,\ldots,x_n) &= \sum_{i\in S}\langle\alpha_i, B(x_i)\rangle + \sum_{1\leq i<j\leq n} B^T(x_i)\beta_{ij}B(x_j) \\
&\leq \sum_{i\in S}\langle\alpha_i, B(x_i)\rangle + (\log 2)\sum_{1\leq i<j\leq n}[\xi^T\beta_{ij}w(x_j) + w^T(x_i)\beta_{ij}\xi] \\
&= \sum_{i\in S}\langle\alpha_i, w(x_i)\rangle + (\log 2)\sum_{1\leq i<j\leq n}\xi^T[\beta_{ij}w(x_j) + \beta_{ij}^T w(x_i)] + R_n \\
&= \sum_{i\in S}\langle\alpha_i + (\log 2)\sum_{j\neq i}\beta_{ij}\xi,\ w(x_i)\rangle + R_n,
\end{aligned}$$

where R_n is a constant, depending only on the parameters α_i. Therefore, up to a constant factor, e^Q is bounded above by a product of n independent beta densities that are well defined, since from (3·4) and (3·5), the exponents of the factors x_i and $(1-x_i)$ are all greater than -1. Hence e^Q is integrable over $(0,1)^n$.

Proof of Theorem 2

Since the $A(x_{V_i}, \theta)$'s are linear functions of the parameters in θ, and e^x is a convex function, the parameter space Θ is a convex set in $\mathbb{R}^q$. Moreover, it is not difficult to see that this linear dependence also implies that the Hessian matrix of $U_n(\theta)$ is everywhere nonnegative-definite. Hence U_n is a convex function of θ.

In the case of a translation-invariant specification, the extremal elements of $\mathcal{G}_s(\theta_0)$ coincide with the stationary ergodic Gibbs measures. Moreover, every element of $\mathcal{G}_s(\theta_0)$ is a convex combination of its ergodic (extremal) elements (Sinai (1982)): consider a probability measure w defined on the set $\mathcal{G}_s^*(\theta_0)$ of ergodic elements of $\mathcal{G}_s(\theta_0)$, then we have

$$\mu_{\theta_0} = \int_{\mathcal{G}_s^*(\theta_0)} \nu \cdot dw(\nu)\ .$$

Define the convergence set of the pseudo-likelihood estimator $\widehat{\theta}_n = \widehat{\theta}_n(x_i, i \in [-n_1, n_1] \times [-n_2, n_2])$:

$$A := \{x \in E^{\mathbb{Z}^2} \; : \; \widehat{\theta}_n \to \theta_0 \;\; \mathrm{as n_1 \to \infty, n_2 \to \infty} \; .$$

If for all ergodic elements ν it holds that $\nu(A) = 1$, we clearly have $\mu_{\theta_0}(A) = 1$. Therefore, without any loss of generality, we can focus on proving the result for μ_{θ_0} ergodic.

Next, by Assumption 1 and the ergodic theorem, the following limit exists almost-surely (μ_{θ_0}),

$$K(\theta) = \lim_n U_n(\theta) = -E_{\mu_{\theta_0}} R(x_{W_0}, \theta).$$

Moreover,

$$K(\theta) = -E_{\mu_{\theta_0}} \left[E_{\mu_{\theta_0}}(R(x_{W_0}, \theta)|x_{V_0}) \right] = E_{\mu_{\theta_0}} \left[D_{KL}(p(\cdot|x_{V_0}, \theta_0), p(\cdot|x_{V_0}, \theta)) \right] \geq 0,$$

where $D_{KL}(P, Q) = \int \log(P/Q) dP$ is the Kullback-Leibler divergence . Furthermore, $K(\theta) = 0$ if and only if $\theta = \theta_0$ under Assumption 3 of the Theorem.

Therefore, standard arguments for convex estimating functions imply that the estimator $\widehat{\theta}_n$ is strongly consistent (see Senoussi (1990), or Guyon (1995) Theorem 3.4.4).

Proof of Proposition 3

We need only to check condition 2 of Theorem 2. Note that the parameter sets A_4 and A_8 are subsets of those defined in Equations (3·7) and (3·9), respectively. The advantage here is that for any $\theta \in A_4$ (or A_8), the parameters $A(x_{V_i}, \theta)$ of the local beta conditional distributions are nonnegative, componentwisely. Consequently, the local contrast function $\{R(x_{W_0}, \theta)\}$ is a continuous function on the compact set $[0, 1]^{|W_0|} \times \Theta$ (previously it could be discontinuous at the boundary 0 and 1). It follows that R is bounded, thus integrable. The conclusion follows.

Acknowledgement. This work originated from several discussions with Xavier Guyon. The authors are also grateful to a referee whose detailed comments have lead to several improvements of the manuscript.

References

D. J. Allcroft, and C. A. Glasbey, 2003. A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. *J. Roy. Statist. Soc.* **C** 52 (4), 487–498.

B. C. Arnold, E. Castillo and J. M. Sarabia, 1999. *Conditional Specification of Statistical Models*. Springer-Verlag, New York.

B. C. Arnold, E. Castillo and J. M. Sarabia, 2001. Conditionally specified distributions: an introduction (with discussion). *Statistical Science* **16** (3), 249-274.

J. Besag, 1974. Spatial interactions and the statistical analysis of lattice systems. *J. Roy. Statist. Soc.* **B** 148, 1-36.

J. Besag, 1977. Efficiency of pseudolikelihood estimation for simple Gaussian fields. *Biometrika* **64** (3), 616–618.

M. S. Bartlett, 1968. A further note on nearest neighbour models. *J. Roy. Statist. Soc.* **A** 131, 579-580.

P. Bouthemy, C. Hardouin, G. Piriou and J. Yao, 2006. Mixed-State auto-models and motion texture modelling. *Journal of Mathematical Imaging and Vision* **25**, (3), 387–402.

F. Comets and M. Janzura, 1998. A central limit theorem for conditionally centered random fields with an application to Markov random fields. *J. Appl. Probab.* **35**, 608–621.

N. Cressie and S. Lele, 1992. New models for Markov random fields. *Adv. Appl. Probab.* **29**, 877-884.

X. Guyon, 1995. *Random Fields on a Network: Modeling, Statistics, and Applications.* Springer-Verlag, New York.

F. Huang and Y. Ogata, 2002. Generalized pseudo-likelihood estimates for Markov random fields on lattice. *Ann. Inst. Statist. Math.* **54**, (1), 1-18.

M. S. Kaiser and N. Cressie, 2000. The construction of multivariate distributions from Markov random fields. *J. Multivariate Anal.* **73**, 199-220.

M. S. Kaiser, N. Cressie and J. Lee, 2002. Spatial mixture models based on exponential family conditional distributions. *Statistica Sinica* **12**, 449-474

J. Lee, M. S. Kaiser and N. Cressie, 2001. Multiway dependence in exponential family conditional distributions. *J. Multivariate Anal.* **79**, 171-190.

Ya. G. Sinai, 1982. *Theory of Phase Transitions: Rigorous Results*. Pergamon Press, Oxford.

R. Senoussi, 1990. Statistique asymptotique presque-sûre de modèles statistiques convexes. *Ann. Inst. Henri Poicaré* **26**, 19-44.

P. Whittle, 1963. Stochastic processes in several dimensions. *Bull. Inst. Statist. Inst.* **40**, 974-994.